Consolidating User-defined Concepts with StYLiD

Aman Shakya
National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku,
Tokyo, Japan 101-8430

shakya_aman@nii.ac.jp

Hideaki Takeda
National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku,
Tokyo, Japan 101-8430

takeda@nii.ac.jp

Vilas Wuwongse
Asian Institute of Technology
Klong Luang, Pathumthani,
Thailand 12120

vw@cs.ait.ac.th

ABSTRACT

Information sharing can be effective with structured data. However, there are several challenges for having structured data on the web. Creating structured concept definitions is difficult and multiple conceptualizations may exist due to different user requirements and preferences. We propose consolidating multiple concept definitions into a unified virtual concept. We have implemented a system called StYLiD to realize this. StYLiD is a social software for sharing a wide variety of structured data. Users can freely define their own structured concepts. Attributes of the multiple concept versions are aligned semi-automatically to provide a unified view. It provides a flexible interface for easy concept definition and data contribution. Popular concepts gradually emerge from the cloud of concepts while concepts evolve incrementally. StYLiD also supports linked data by interlinking data instances including external resources like Wikipedia.

Categories and Subject Descriptors

H.4.m [**Information Systems**]: Miscellaneous

General Terms

Design

Keywords

Structured data, concept consolidation, social Semantic Web, information sharing, linked data.

1. INTRODUCTION

People want to share a wide variety of information on the web. Different types of data can be modeled by structuring them. The semantics of structured data facilitates automated processing. Information sharing, searching and browsing become more effective. Interoperability and integration of data from various sources becomes possible with standard or compatible formats.

However, there are several challenges for sharing structured data on the web. There is a long tail of information domains[1] for which people have information to share but not many information systems or ontologies to cover the wide variety of information [2]. Developing individual solutions is infeasible because creating new ontologies or information systems is difficult. It is not easy to define concepts adequately. Moreover, people conceptualize the same thing in different ways based on contexts, requirements or preferences. Creating ontologies should be a widely collaborative and incremental process[2]. However, to

have mass participation, systems should be easy to understand and use. Unlike ontologies, social software has proven to be very successful in that. Thus, the combination of social software and Semantic Web technologies has attracted significant attention.

In an attempt to address these issues, we propose a social software to enable general users to freely define their own conceptualizations and share structured data based on that. We propose allowing users to have multiple conceptualizations and, at the same time, consolidating them into a single virtual conceptualization. The consolidation of multiple user-defined concepts serves as a new way of building up richer concept definitions collaboratively. We have implemented a system called StYLiD[1] (an acronym for Structure Your own Linked Data) to realize our approach. Multiple definitions are consolidated by semi-automatic alignment of concept attributes. Popular concepts can gradually emerge out by usage in the same way as folksonomies. Besides, StYLiD also produces linked data by interlinking internal and external resources. The system integrates various aspects of social software and Semantic Web technologies into a synergetic whole.
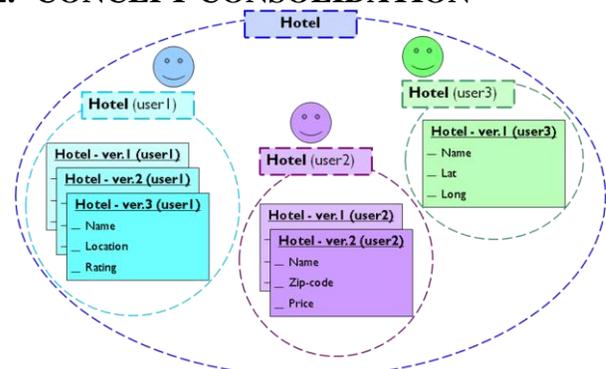
2. CONCEPT CONSOLIDATION

Figure 1: Concept consolidation.

Different users may have multiple definitions for the same concept as illustrated in Fig. 1. Even the same user may have multiple versions. The system groups them and consolidates them into a single virtual concept combining the features of the individual definitions. Our approach for consolidation is based on the Global-as-View approach for data integration where a global schema is defined in terms of the source schemas[3].

---

[1] http://www.stylid.org/

3. THE STYLID PLATFORM

3.1 Sharing Data with User-Defined Concepts

Suppose a user wants to share some information about a talk program. He may define his own "talk" concept on the fly with a list of attributes like topic, speaker, date, venue, etc. Instead of defining from scratch, he may also modify an existing concept to create his own version. Concepts can evolve incrementally. Once a concept has been defined, he (or any other user) may easily start posting data instances. He may easily share them in his community using social software features.

It is difficult to think of all attributes and possible values while defining concepts. Further, we may not have perfect data for entry. We try to avoid these difficulties by allowing flexible and relaxed definitions. The range of values does not impose constraints. Rather the suggestive range assists the user to pick up instances. The system accepts both literal values or resource URIs and single or multiple values for any attribute.

3.2 Consolidation of User Defined Concepts

All user-defined concepts are visualized in a *Concept Cloud*. Clicking on a concept shows all its instances while hovering on it shows its details. Popular definitions gradually emerge bigger from the cloud as more instances are contributed. Different users may define multiple versions of the same concept. Concepts with the same name are grouped together forming a virtual consolidated concept. This can be expanded into a sub-cloud of all the versions, labeled with the creator name and version number. In the sub-cloud, versions defined by the same user are subgrouped together. The sizes of all versions in the sub-cloud add up to form the size of the consolidated concept. Clicking on the consolidated concept shows all instances of all its versions. We can also list instances of the versions defined by a single user by clicking on the user name.

The concepts in a consolidated group can be aligned to produce a unified view. When instances of a consolidated concept are viewed as a single table, the system automatically suggests alignments between the attributes. The Alignment API[2] with its WordNet extension has been used. The user has to complete the process by adding or modifying mappings not correctly suggested. Once completed by a user, the alignments are saved for all. Thus, both machine intelligence and human intelligence are used in getting the concepts aligned. Each set of aligned attributes is mapped to a single consolidated attribute. The user can create a unified view, customized according to his need, and view heterogeneous data in a uniform table. The table can be sorted by any field.

3.3 Creating Linked Data

The system creates linked data by providing unique dereferenceable URIs for each concept, attribute and instance. Data instances can be linked by entering resource URIs as attribute values. Any external URI may be directly linked as an attribute value. The system provides support to link to Wikipedia contents. Unlike DBpedia, Wikipedia is well understood by general people and user-friendly. Transparent to the user, the system converts Wikipedia URLs into DBpedia URIs.

3.4 Querying and Using Structured Data

A structured search interface is provided to retrieve instances of a concept by specifying attribute, value pairs. If the search is done over a consolidated concept, the query is unfolded to the aligned attributes of all versions of the concept. It also provides a SPARQL query interface. The user can navigate through linked data entries. Structured data is also embedded in the HTML posts using RDFa. Many useful RDFa tools are becoming available. Users may even code scripts with the Operator[3] extension to create useful operations for different types of data.

4. RELATED WORK

Freebase[4] allows users to define their own structured types. However, these are kept separate and not consolidated or related in any way. It is also difficult to contribute types and instances because of strict constraints. Exhibit[1] enables ordinary users to create web pages with structured data. However, authoring such pages would be cumbersome. myOntology[2] uses wikis for collaborative lightweight ontology building by enabling general users. Freebase, myOntology and semantic wikis are all based on wikis which assume that each resource can only have a single prominent version. However, in practice, multiple conceptualizations may exist.

5. CONLUSION AND ONGOING WORK

We proposed an approach for consolidating multiple user-defined conceptualizations. It facilitates collaborative creation of knowledge while meeting individual requirements of different people. We proposed StYLiD as a social software for sharing any type of structured data. Ontologies can be formed as a by-product of usual information sharing activities in the community.

We are working on grouping similar concepts and computing relations between them. This can be done considering the structure definitions, and using lexical resources like WordNet. Better alignment techniques may be employed. We should also reuse existing vocabularies or map concept definitions to them.

6. REFERENCES

[1] Huynh, D., Karger, D. and Miller, R. 2007. Exhibit: lightweight structured data publishing. In Proceedings of the 16th international conference on World Wide Web. ACM Press, New York, 737–746.

[2] Siorpaes, K. and Hepp, M. 2007. myOntology: The marriage of ontology engineering and collective intelligence. In Bridging the Gap between Semantic Web and Web 2.0 (SemNet 2007). 127–138

[3] Lenzerini, M. 2002. Data integration: A theoretical perspective. In Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems. 233–246

---

[2] http://alignapi.gforge.inria.fr/

[3] http://www.kaply.com/weblog/operator/

[4] http://www.freebase.com/